Proposed Panel Conclusions and Recommendations for the Isolated Rabbit Eye (IRE) Test Method

Expert Panel Meeting

National Institutes of Health
Natcher Conference Center
Bethesda, Maryland
January 11-12, 2005

IRE Test Method Primary Reviewers

- Jim Freeman, Ph.D., D.A.B.T., *ExxonMobil, USA*
- Sally Atherton, PhD, *Med College Georgia, USA*
- David Lovell, Ph.D., *University of Surrey, United Kingdom* (Biostatistician)
- Yasuo Ohno, Ph.D., D.J.S.T., *NIHS*, Japan
- Horst Spielmann, Dr. Med., *ZEBET at the BfR, Germany*
- Peter Theran, V.M.D., D.A.C.V.I.M., *Massachusetts SPCA, CA*

BRD Section 1.0:
IRE Test Method Rationale

1.1 Scientific Basis for the IRE Test Method

Concur with description in BRD

Recommendations:

- **Add discussion of cellular mechanisms of corrosion and severe irritation (e.g., necrosis, apoptosis) and relevance to *in vitro* testing. (ICE, HET-CAM)**

- **Discuss the role of responsive inflammatory cells in isolated rabbit eyes. (ICE, CAM in HET-CAM)**

- **Update BRD to discuss the basis of the assay as a correlation of descriptive observations of toxicity, rather than mechanistic (e.g., corneas may appear similar, but damaged by different mechanisms)**

1.2 Regulatory Rationale and Applicability

Thoroughly covered in BRD

Comments:

IRE, as described in the protocol,:

- Does not account for effects on the iris and conjunctiva
- may not account for the reversibility of corneal effects
- Does not account for systemic effects
- Does not identify slow-acting irritants (on the order of days)

Recommendations:

- Consider use of microscopy or histopathology to improve sensitivity and scope
  - ➢ early markers of effect
  - ➢ identify transient vs progressive changes

BRD Section 2.0:
IRE Test Method Protocol Components

General comments on Section 2.0:

- Limited data set derived from recommended protocol
- Recommended protocol enhancements improve accuracy
- Standardized protocol has not been directly assessed across laboratories
- The BRD should identify the decision criteria (Prediction Model) for identifying ocular corrosives and severe irritants and discuss rationale for development
- Identify acceptable positive or negative controls or reference substances from validation reference substance list

2.1 Test Method Components for the Recommended Version of the Protocol (Appendix A)

Recommendations:

- Appropriate sources of rabbit eyes needs to be defined; specify acceptable strain(s) of rabbit; define acceptable storage/transport conditions of eyes and evaluation prior to shipping
- Use corneal opacity and area, corneal thickness, corneal swelling, fluorescein penetration, and epithelial integrity as endpoints as described
- Identification of reference substances that are part of the performance standards developed for the validated test method
- Collect and store data using GLP compliant procedures
- Clarify in the BRD orientation of eye (vertical or horizontal) during TM application
- Consider confocal microscopy or histopathology to detect changes at the cellular level and quantitation of fluorescein observation (i.e., counting pixels)
- Use descriptive statistics based on individual scores
- Identify the Prediction Model more clearly in BRD

2.2 Basis for Selection of the Test Method System

Well presented in BRD

2.3 Proprietary Components

Not applicable; do not believe the apparatus is proprietary

2.4 Number of Replicate and/or Repeat Experiments for Each Test

Concur with BRD

- Statistical methods are appropriate and the conclusions on reliability are basically sound

2.5 Study Acceptance Criteria

Concur with BRD

- Acceptance determined by appropriate response of reference controls.

2.6 Basis for any Modifications to the Original Test Method Protocol

Adequately described in BRD

- Any further additions must be backed by specific rationale

2.7 Adequacy of the Recommended Standardized Protocol

Adequately covered in BRD

- Consider inclusion of histopathology as described in Section 1.2.2
- Update recommended protocol with reference chemical list developed by Expert Panel

BRD Section 3.0: Substances Used for Previous Validation Studies of the IRE Test Method

3.1 Types & Numbers of Substances/Products Used for Prior Validation Studies

Adequately described in BRD

- Further optimization/validation requires use of reference substances recommended by Expert Panel Reference Substances Subgroup

3.2 Coding Procedures for Test Substances and Quality of IRE Test Method Data

Adequately described in BRD

- Coding procedures appear to have been adequate and did not introduce bias

BRD Section 4.0: *In Vivo* Reference Data Used for an Assessment of Test Method Accuracy

4.1 *In Vivo* Rabbit Eye Test Method Protocols Used to Generate Reference Data

Appropriately described in BRD

- Panel notes that Draize test is basically unchanged (no improvements) for decades
- Perhaps Draize test can be improved *vis a vis* ongoing / future testing using technology being considered for *in vitro* studies (e.g., fluorescein, slit lamp, etc)

4.2 Interpretation of *In Vivo* Test Method Results for Cited Studies

Interpretation of results were correct in BRD

- Question raised regarding adequacy of using regulatory classification systems for evaluating *in vitro* methods and suitability for chemical or product class evaluations - rewording needed

4.3 Data Quality for Test Substances When Original Study Records Are or Are Not Available

Acceptable as written in BRD

- If evaluation of results can be made and the quality of the study is adequate, lack of original records does not raise concerns about a study

4.4 Data Quality With Respect to Extent of GLP Compliance

Acceptable as written in BRD

- If work is performed in well established laboratories, no distinction between GLP compliant *versus* non-GLP compliant studies is required
- Lack of GLP compliance is not a sufficient criterion for exclusion of data for evaluation of performance

4.5 Availability of Relevant Human Ocular Toxicity Information

- **There needs to be greater effort to obtain and consider information on human topical ocular chemical injury.**
- **Very limited human ocular exposure data is available and comparison with respect to dose received and exposure time could be difficult to quantitate.**
- **No scoring or time course data would be available for comparison to an *in vivo* or *in vitro* test method**

4.6 Accuracy and Reliability of the *In Vivo* Rabbit Eye Test

- Need more discussion of variability of *in vivo* data
- Is rabbit data consistent with known human data?
- Are inconsistencies due to failure of *in vitro* method or to misclassification of the single *in vivo* result?
- Any optimization and validation studies should use existing animal data; if available.
- Additional animal studies should only be conducted if important data gaps are identified and such studies should be carefully designed to maximize the amount of pathophysiological information obtained (e.g., wound healing).
- Minority opinion – no animal testing for this purpose.

BRD Section 5.0: IRE Test Method Data and Results

5.1 IRE Test Method Protocols Used to Generate Each Set of Data

Adequately described in BRD

- Recommended protocol includes additional parameters that enhance accuracy (Guerriero et al. 2004)

5.2 Other Comparative IRE - *In Vivo* Rabbit Eye Test Data Not Considered in the BRD

Adequately described in BRD

- There are no additional data sets produced with the IRE test method

5.3 Statistical and Nonstatistical Approaches Used to Evaluate the Resulting IRE Data

Adequately described in BRD

- Statistical methods were limited but appropriate for descriptive toxicity data and conclusions on reliability basically sound

5.4 Use of Coded Substances, Blind Studies, and GLP Guidelines for Cited Studies

Adequately described in BRD

- Documentation of data quality is adequate; studies using recommended protocol were conducted according to “the spirit of GLPs”

5.5 "Lot-to-Lot" Consistency of Test Substances and Timeframe of Studies

Adequately described in BRD

- **Referenced IRE studies were independent efforts**
- **"Lot-to-lot" consistency was controlled and described in 3 of the 4 studies; not described in the fourth**
- **Stability of chemicals over each study's timeframe was not discussed in BRD**

BRD Section 6.0: IRE Test Method Accuracy

a. The closeness of agreement between a test method result and an accepted reference value.

b. The proportion of correct outcomes of a test method

6.1 Accuracy Evaluation of the IRE Test Method for Identifying Ocular Corrosives and Severe Irritants as Defined by the EPA (1996), the EU (2001), and the GHS (2003)

Adequately described in BRD

- **Accuracy results summarized in Section 6.1 (Tables 6.1, 6.2, and 6.3) of BRD provide correct overview of performance as reported in the studies, as well as the discordant results; accuracy appears to be (small n) improved with the recommended method, resulting in a false negative rate of 0% and a false positive rate of ~33%**
- **Draize variability (Weil and Scala, 1971; Balls et al. 1995; Spielmann, 1997) must be included in discussion**
- **Weakness with lack of a common protocol for all studies evaluated**
- **Encouraging that accuracy improved in Guerriero et al (2004) dataset upon which the recommended protocol is based**

6.2 Strengths & Limitations of the Test Method, Including Those Applicable to Specific Chemical Classes or to Certain Physicochemical Properties

Adequately described in BRD

Recommendations:

- **Assure correct temporal sequence of studies is clearly written in BRD (e.g., lines 414-415)**
- **Add *in vivo* and *in vitro* source/reference/author information to both Tables in Section 6.3 (Tables 6.4 and 6.5) and identify which *in vitro* data sets were used to calculate the IRE classifications**
- **Additional data is needed to accurately assess overprediction (e.g., alcohols) using a larger set of appropriate reference substances**

6.3 Issues of Data Interpretation

Recommendations:

- **Differences in reproducibility/variability of the Draize test must be taken into account when comparing predictive value of *in vitro* alternatives**
- **Other relevant information (e.g., by a weight of evidence approach) may clarify IRE test performance**
- **Also recognize that the variability of the Draize test for corrosive or severe irritants is lower than what occurs for moderate irritants**

BRD Section 7.0:
IRE Test Method Reliability (Repeatability/Reproducibility)

A measure of the degree to which a test method can be performed reproducibly within and among laboratories over time.

Section 7.0:
IRE Test Method Reliability
(Repeatability/Reproducibility)

General Remarks:

- Incorporate information from Bland and Altman (1986) which discusses comparison of methods with poor reproducibility
- Incorporate information from ECVAM Skin Irritation prevalidation study on repeatability and reproducibility analysis
- Incorporate information on detailed variability analysis (Dr. Hofmann) comparing SD and CV for two skin irritation models
- Develop and implement strategy to evaluate reliability in any future optimization/validation testing

7.1 Selection Rationale for the Substances Used to Evaluate Test Method Reliability

Adequately covered in BRD

- Described in appropriate detail in the relevant publications

7.2 Analyses & Conclusions Regarding Intralaboratory Repeatability and Intra- & Inter-laboratory Reproducibility

Concur with BRD

- No data was provided for multiple studies from a single laboratory; neither intralaboratory repeatability nor reproducibility could be assessed
- Quantitative interlaboratory reproducibility was assessed in 2 of the 4 studies which used slightly different protocols

Recommendation

- Reproducibility analyses should be conducted from studies using the recommended protocol and list of reference chemicals

7.3 Availability of Historical Negative & Positive Control Data

Appropriately covered in BRD

- Positive controls have not been consistently run
- Future studies, however, should track control information

7.4 Effect of Minor Protocol Changes to Recommended Test Method Protocol and Transferability of Test Method

Concur with BRD

Comments and Recommendations:

- There are no impediments to minor protocol changes or transferability of the IRE test method
- May be useful to contrast results developed using SafePharm recommended protocol *vs* earlier renditions; good agreement across the board with *in vivo* data would suggest that existing data from all protocols could be used as validation data
- Any differences in protocols used for future studies should be specifically justified

BRD Section 8.0: IRE Test Method Data Quality

8.1 Extent of Adherence to GLP Guidelines and Use of Coded Chemicals

Concur with BRD

- Lack of GLP compliance *per se* is not an exclusion criterion
- Although not all studies were considered GLP compliant, the reviewed data appeared to be of satisfactory quality

8.2 Data Quality Audits

Appropriately covered in BRD

- Verification of accuracy of data against original data records was beyond scope of IRE assessment

8.3 Impact of Deviations from GLP Guidelines

Appropriately covered in BRD

- Noncompliance with GLP was not a mandatory exclusion criteria
- All laboratories performing the studies were reputable

8.4 Availability of Laboratory Notebooks or Other Records for an Independent Audit

Appropriately covered in BRD

- **Original raw *in vitro* data for all studies was not available for review; availability and review of raw data would improve confidence in the data**
- **Doing retrospective GLP-like audits may not be needed and would be difficult**

BRD Section 9.0: Other Scientific Reports and Reviews

9.1 Adequacy and Completeness of Relevant Data Identified in Other Published or Unpublished IRE Studies

Appropriately covered in IRE BRD

- Submitted P&G ExRET and LVET data not readily comparable to other studies for regulatory classification; excluded from overall analysis
- Included reviews of all relevant published IRE studies

9.2 Adequacy and Completeness of the Conclusions Published in Independent Peer Reviewed Reports or Other Independent Scientific Reviews

Appropriately covered in IRE BRD

- Conclusions reached from report summaries were adequate and complete

9.3 Approaches that can be Used to Expedite the Process for Obtaining Additional In-House Data from the Private Sector

Appropriately covered in IRE BRD

- FR Notice (V69,N7,13859-13861) sent 3/24/04 requesting IRE test method data
- Authors contacted to request original IRE data and *in vivo* reference data

BRD Section 10.0: Animal Welfare Considerations (Refinement, Reduction, Replacement)

10.1 Extent to Which the Test Method Will Refine, Reduce or Replace Animal Use

Appropriately covered in BRD

Comments and Recommendations:

- Determine actual availability of rabbits
- Rabbits should not be raised and killed specifically for use in this test.
- NICEATM should determine if the current policy of which, if any, US regulatory agencies would not accept the use of eyes from rabbits used for other scientific purposes.
- The BRD should review the availability of rabbit eyes to US labs.
- Availability of eyes from an abattoir may be a factor for further development of this test method
- Method could be a partial replacement if eyes are available

BRD Section 11.0: Practical Considerations

11.1 Adequacy and Completeness of Test Method Transferability

Appropriately covered in BRD

- Transferability appears readily achievable
- A training video and other visual media on the technical aspects of the assay is recommended (place in all)
- Training approaches in the application of this test method should be developed/implemented (place in all)

11.2 Adequacy and Completeness of Test Method Training

Appropriately covered in BRD

- Experienced personnel should provide training

11.3 Adequacy and Completeness of Information on *In Vitro* and *In Vivo* Cost

Adequately described

- **Information from one UK lab quotes the cost for the IRE with controls at $1074 and the cost of the *In vivo* Rabbit Eye Test (n=3) at $969-1709 (depending upon length of study).**
- **Costs in the US can be expected to be greater.**

11.4 Adequacy and Completeness of Information on the Amount of Time Needed to Conduct a Study

Adequately described

- ***The in vivo* test may require up to 21 days; however, it is recognized that a corrosive or severe irritant may be detected within a few hours using a single rabbit.**
- **The recommended IRE test method can be completed, from the onset of treatment, in approximately 4 hours**

BRD Section 12.0: Proposed IRE Test Method Recommendations

12.1 Proposed Version of the IRE Test Method

Concur with BRD

Comments and Recommendations:

- Recommended version of IRE protocol has only been conducted in one laboratory
- Limited data were generated using the recommended protocol
- Of the methods presented in the BRD, the panel agrees that the most appropriate version of the protocol was selected and adding fluorescein staining and epithelial integrity to the measurement of corneal thickness and opacity adds to the accuracy of the test

12.2 Proposed Standardized IRE Test Method Protocol (1)

- Appropriate source of rabbit eyes needs to be defined.
- The current policy of some US regulatory agencies (e.g., EPA) in regard to the use of eyes from rabbits used for other scientific studies should be reviewed and updated.
- Rabbits should not be raised and killed specifically for use in this test
- Prediction model is used, but needs to be identified more clearly in BRD
- Defined positive or negative controls or reference substances should be added based on the recommended Reference Substances list provided by the Expert Panel

12.2 Proposed Standardized IRE Test Method Protocol (2)

- A standardized scoring scheme for histopathology should be defined using the formal language of pathology to describe any effects
- The appropriate circumstances under which histopathology would be warranted should be more clearly defined
- To maximize the likelihood of obtaining reproducible results, reference photographs for all subjective endpoints (i.e., corneal opacity, fluorescein penetration, and histopathology) should be developed to aid training and transferability

12.3 Proposed IRE Optimization and Validation Studies (1)

- Recommended IRE method appears to be capable of identifying severe irritants/corrosives in a tiered testing strategy; however, database is so small (n=36 classifiable to GHS), and there is a lack of any data on reproducibility; thus, a decision can not be made with the currently available data and more data needs to be considered before an appropriate evaluation of the IRE test for classification can be conducted
- Existing data with recommended protocol indicates a relatively high false positive rate; false negative rate of 0 (n=12) is encouraging
- Appropriate sources of rabbit eyes must be identified if further optimization and validation is to proceed

12.3 Proposed IRE Optimization and Validation Studies (2)

- If further work on IRE entails, then optimize to reduce false positive rate without unacceptably increasing false negative rate
- A high quality database of *in vivo* and *in vitro* data of reference substances should be established from existing literature and new data
- Employment of statistical methods (e.g., discriminant analysis) may be used to derive a more general Prediction Model for the IRE; the PM must be optimized with the existing data

12.3 Proposed IRE Optimization and Validation Studies (3)

- Any optimization and validation studies should use existing animal data; if available.
- Additional animal studies should only be conducted if important data gaps are identified and such studies should be carefully designed to maximize the amount of pathophysiological information obtained (e.g., wound healing)
- Minority opinion –sufficient data should be available so no animal testing for this purpose – Dr. Stephens

12.3 Proposed IRE Optimization and Validation Studies (4)

- **NICEATM/ICCVAM should facilitate the development of a histopathology scoring system for corneal damage (with visual aids)**